

Docket No. 10806-156

PATENT

CERTIFICATE OF MAILING

I hereby certify that this paper is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Box Missing Parts; Commissioner of Patents, Washington, DC 20231 on March 12, 2002.

Laureen E. Maens

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

Applicant: Anneli Attersand : Paper No.:
Serial No.: 10/029,359 : Group Art Unit: 1645
Filing Date: December 21, 2001 : Examiner:
For: **Protein Cluster II**

TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Box Missing Parts
Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

Submitted herewith is a certified copy of Swedish priority application No. 0004828-0 filed December 22, 2000. The priority document is in the English language. It is believed that this satisfies the requirements of 35 U.S.C. §119.

Respectfully submitted,

By Holly D. Kozlowski
Holly D. Kozlowski, Reg. No. 30,468
Attorney for Applicant

DINSMORE & SHOHL LLP
1900 Chemed Center
255 East Fifth Street
Cincinnati, Ohio 45202
(513) 977-8568

PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen



**COPY OF PAPERS
ORIGINALLY FILED**

**Intyg
Certificate**

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.



(71) Sökande *Pharmacia AB, Stockholm SE*
Applicant (s)

(21) Patentansökningsnummer *0004828-0*
Patent application number

(86) Ingivningsdatum *2000-12-22*
Date of filing

Stockholm, 2001-12-17

*För Patent- och registreringsverket
For the Patent- and Registration Office*

Christina Vängborg
Christina Vängborg

Avgift
Fee *170:-*

PROTEIN CLUSTER II

TECHNICAL FIELD

The present invention relates to the identification of a human gene family expressed in metabolically relevant tissues. The genes encode a group polypeptides referred to as "Protein Cluster II" which are predicted to be useful in the diagnosis of metabolic diseases, such as obesity and diabetes, as well as in the identification of agents useful in the treatment of the said diseases.

BACKGROUND ART

Metabolic diseases are defined as any of the diseases or disorders that disrupt normal metabolism. They may arise from nutritional deficiencies; in connection with diseases of the endocrine system, the liver, or the kidneys; or as a result of genetic defects. Metabolic diseases are conditions caused by an abnormality in one or more of the chemical reactions essential to producing energy, to regenerating cellular constituents, or to eliminating unneeded products arising from these processes. Depending on which metabolic pathway is involved, a single defective chemical reaction may produce consequences that are narrow, involving a single body function, or broad, affecting many organs and systems.

One of the major hormones that influence metabolism is insulin, which is synthesized in the beta cells of the islets of Langerhans of the pancreas. Insulin primarily regulates the direction of metabolism, shifting many processes toward the storage of substrates and away from their degradation. Insulin acts to increase the transport of glucose and amino acids as well as key minerals such as potassium, magnesium, and phosphate from the blood into cells. It also regulates a variety of enzymatic reactions within the cells, all of which have a common overall direction, namely the synthesis of large molecules from small units. A deficiency in the action of insulin (diabetes mellitus) causes severe impairment in (i) the storage of glucose in the form of glycogen and the oxidation of

glucose for energy; (ii) the synthesis and storage of fat from fatty acids and their precursors and the completion of fatty-acid oxidation; and (iii) the synthesis of proteins from amino acids.

There are two varieties of diabetes. Type I is insulin-dependent diabetes mellitus (IDDM), for which insulin injection is required; it was formerly referred to as juvenile onset diabetes. In this type, insulin is not secreted by the pancreas and hence must be taken by injection. Type II, non-insulin-dependent diabetes mellitus (NIDDM) may be controlled by dietary restriction. It derives from insufficient pancreatic insulin secretion and tissue resistance to secreted insulin, which is complicated by subtle changes in the secretion of insulin by the beta cells. Despite their former classifications as juvenile or adult, either type can occur at any age; NIDDM, however, is the most common type, accounting for 90 percent of all diabetes. While the exact causes of diabetes remain obscure, it is evident that NIDDM is linked to heredity and obesity. There is clearly a genetic predisposition to NIDDM diabetes in those who become overweight or obese.

Obesity is usually defined in terms of the body mass index (BMI), i.e. weight (in kilograms) divided by the square of the height (in meters). Weight is regulated with great precision. Regulation of body weight is believed to occur not only in persons of normal weight but also among many obese persons, in whom obesity is attributed to an elevation in the set point around which weight is regulated. The determinants of obesity can be divided into genetic, environmental, and regulatory.

Recent discoveries have helped explain how genes may determine obesity and how they may influence the regulation of body weight. For example, mutations in the *ob* gene have led to massive obesity in mice. Cloning the *ob* gene led to the identification of leptin, a protein coded by this gene; leptin is produced in adipose tissue cells and acts to control body fat. The existence of leptin supports the idea that body weight is regulated, because leptin serves as a signal between adipose tissue and the areas of the brain that control energy metabolism, which influences body weight.

Metabolic diseases like diabetes and obesity are clinically and genetically heterogeneous disorders. Recent advances in molecular genetics have led to the recognition of genes involved in IDDM and in some subtypes of NIDDM, including maturity-onset diabetes of the young (MODY) (Velho & Froguel (1997) Diabetes Metab. 23 Suppl 2:34-37). However, several IDDM susceptibility genes have not yet been identified, and very little is known about genes contributing to common forms of NIDDM. Studies of candidate genes and of genes mapped in animal models of IDDM or NIDDM, as well as whole genome scanning of diabetic families from different populations, should allow the identification of most diabetes susceptibility genes and of the molecular targets for new potential drugs. The identification of genes involved in metabolic disorders will thus contribute to the development of novel predictive and therapeutic approaches.

DESCRIPTION OF THE INVENTION

According to the present invention, a family of genes and encoded homologous proteins (hereinafter referred to as "Protein Cluster II") has been identified. Consequently, the present invention provides an isolated nucleic acid molecule selected from:

- (a) nucleic acid molecules comprising a nucleotide sequence as shown in SEQ ID NO: 1, or 3;
- (b) nucleic acid molecules comprising a nucleotide sequence capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence complementary to the polypeptide coding region of a nucleic acid molecule as defined in (a); and
- (c) nucleic acid molecules comprising a nucleic acid sequence which is degenerate as a result of the genetic code to a nucleotide sequence as defined in (a) or (b).

The nucleic acid molecules according to the present invention includes cDNA, chemically synthesized DNA, DNA isolated by PCR, genomic DNA, and combinations thereof. RNA transcribed from DNA is also encompassed by the present invention.

The term "stringent hybridization conditions" is known in the art from standard protocols (e.g. Ausubel et al., *supra*) and could be understood as e.g. hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at +65°C, and washing in 0.1xSSC / 0.1% SDS at +68°C.

In a preferred form of the invention, the said nucleic acid molecule has a nucleotide sequence identical with SEQ ID NOS: 1 or 3 of the Sequence Listing. However, the nucleic acid molecule according to the invention is not to be limited strictly to the sequence shown as SEQ ID NOS: 1 or 3. Rather the invention encompasses nucleic acid molecules carrying modifications like substitutions, small deletions, insertions or inversions, which nevertheless encode proteins having substantially the features of the Protein Cluster II polypeptide according to the invention. Included in the invention are consequently nucleic acid molecules, the nucleotide sequence of which is at least 90% homologous, preferably at least 95% homologous, with the nucleotide sequence shown as SEQ ID NOS: 1 or 3 in the Sequence Listing.

Included in the invention is also a nucleic acid molecule which nucleotide sequence is degenerate, because of the genetic code, to the nucleotide sequence shown as SEQ ID NO: 1 or 3. A sequential grouping of three nucleotides, a "codon", codes for one amino acid. Since there are 64 possible codons, but only 20 natural amino acids, most amino acids are coded for by more than one codon. This natural "degeneracy", or "redundancy", of the genetic code is well known in the art. It will thus be appreciated that the nucleotide sequence shown in the Sequence Listing is only an example within a large but definite group of sequences which will encode the Protein Cluster II polypeptide.

The nucleic acid molecules according to the invention have numerous applications in techniques known to those skilled in the art of molecular biology. These techniques include their use as hybridization probes, for chromosome and gene mapping, in PCR technologies, in the production of sense or antisense nucleic acids, in screening for new therapeutic molecules, etc.

More specifically, the sequence information provided by the invention makes possible large-scale expression of the encoded polypeptides by techniques well known in the art. Nucleic acid molecules of the invention also permit identification and isolation of nucleic acid molecules encoding related polypeptides, such as human allelic variants and species homologues, by well-known techniques including Southern and/or Northern hybridization, and PCR. Knowledge of the sequence of a human DNA also makes possible, through use of Southern hybridization or PCR, the identification of genomic DNA sequences encoding the proteins in Cluster II, expression control regulatory sequences such as promoters, operators, enhancers, repressors, and the like. Nucleic acid molecules of the invention are also useful in hybridization assays to detect the capacity of cells to express the proteins in Cluster II. Nucleic acid molecules of the invention may also provide a basis for diagnostic methods useful for identifying a genetic alteration(s) in a locus that underlies a disease state or states, which information is useful both for diagnosis and for selection of therapeutic strategies.

In a further aspect, the invention provides an isolated polypeptide encoded by the nucleic acid molecule as defined above. In a preferred form, the said polypeptide has an amino acid sequence according to SEQ ID NO: 2 or 4 of the Sequence Listing. However, the polypeptide according to the invention is not to be limited strictly to a polypeptide with an amino acid sequence identical with SEQ ID NO: 2 or 4 in the Sequence Listing. Rather the invention encompasses polypeptides carrying modifications like substitutions, small deletions, insertions or inversions, which polypeptides nevertheless have substantially the features of the Protein Cluster II polypeptide. Included in the invention are consequently polypeptides, the amino acid sequence of which is at least 90% homologous, preferably at least 95% homologous, with the amino acid sequence shown as SEQ ID NO: 2 or 4 in the Sequence Listing.

In a further aspect, the invention provides a vector harboring the nucleic acid molecule as defined above. The said vector can e.g. be a replicable expression vector, which carries and is capable of mediating the expression of a DNA molecule according to the invention. In the present context the term "replicable" means that the vector is able to replicate in a given type of host cell into which it has been introduced. Examples of

vectors are viruses such as bacteriophages, cosmids, plasmids and other recombination vectors. Nucleic acid molecules are inserted into vector genomes by methods well known in the art.

Included in the invention is also a cultured host cell harboring a vector according to the invention. Such a host cell can be a prokaryotic cell, a unicellular eukaryotic cell or a cell derived from a multicellular organism. The host cell can thus e.g. be a bacterial cell such as an *E. coli* cell; a cell from a yeast such as *Saccharomyces cerevisiae* or *Pichia pastoris*, or a mammalian cell. The methods employed to effect introduction of the vector into the host cell are standard methods well known to a person familiar with recombinant DNA methods.

In yet another aspect, the invention provides a process for production of a polypeptide, comprising culturing a host cell, according to the invention, under conditions whereby said polypeptide is produced, and recovering said polypeptide. The medium used to grow the cells may be any conventional medium suitable for the purpose. A suitable vector may be any of the vectors described above, and an appropriate host cell may be any of the cell types listed above. The methods employed to construct the vector and effect introduction thereof into the host cell may be any methods known for such purposes within the field of recombinant DNA. The recombinant polypeptide expressed by the cells may be secreted, i.e. exported through the cell membrane, dependent on the type of cell and the composition of the vector.

In a further aspect, the invention provides a method for identifying an agent capable of modulating a nucleic acid molecule according to the invention, comprising

- (i) providing a cell comprising the said nucleic acid molecule;
- (ii) contacting said cell with a candidate agent; and
- (iii) monitoring said cell for an effect that is not present in the absence of said candidate agent.

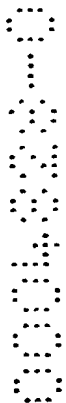
For screening purposes, appropriate host cells can be transformed with a vector having a reporter gene under the control of the nucleic acid molecule according to this invention.

The expression of the reporter gene can be measured in the presence or absence of an agent with known activity (i.e. a standard agent) or putative activity (i.e. a "test agent" or "candidate agent"). A change in the level of expression of the reporter gene in the presence of the test agent is compared with that effected by the standard agent. In this way, active agents are identified and their relative potency in this assay determined.

A transfection assay can be a particularly useful screening assay for identifying an effective agent. In a transfection assay, a nucleic acid containing a gene such as a reporter gene that is operably linked to a nucleic acid molecule according to the invention, is transfected into the desired cell type. A test level of reporter gene expression is assayed in the presence of a candidate agent and compared to a control level of expression. An effective agent is identified as an agent that results in a test level of expression that is different than a control level of reporter gene expression, which is the level of expression determined in the absence of the agent. Methods for transfecting cells and a variety of convenient reporter genes are well known in the art (see, for example, Goeddel (ed.), *Methods Enzymol.*, Vol. 185, San Diego: Academic Press, Inc. (1990); see also Sambrook, *supra*).

Throughout this description the terms "standard protocols" and "standard procedures", when used in the context of molecular biology techniques, are to be understood as protocols and procedures found in an ordinary laboratory manual such as: *Current Protocols in Molecular Biology*, editors F. Ausubel et al., John Wiley and Sons, Inc. 1994, or Sambrook, J., Fritsch, E.F. and Maniatis, T., *Molecular Cloning: A laboratory manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1989.

Additional features of the invention will be apparent from the following Examples. Examples 1 to 3 are actual, while Examples 4 to 9 are prophetic.



EXAMPLES

EXAMPLE 1: Identification of protein clusters

A family of homologous proteins (hereinafter referred to as "Protein Cluster II") was identified by an "all-versus-all" BLAST procedure using all *Caenorhabditis elegans* proteins in the Wormpep20 database release (http://www.sanger.ac.uk/Projects/C_elegans/wormpep/index.shtml). The Wormpep database contains the predicted proteins from the *C. elegans* genome sequencing project, carried out jointly by the Sanger Centre in Cambridge, UK and the Genome Sequencing Center in St. Louis, USA. A number of 18,940 proteins were retrieved from Wormpep20. The proteins were used in a Smith-Waterman clustering procedure to group together proteins of similarity (Smith T.F. & Waterman M.S. (1981) *Identification of common molecular subsequences*. J. Mol. Biol. 147(1): 195-197; Pearson WR. (1991) *Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms*. Genomics 11: 635-650; Olsen et al. (1999) *Optimizing Smith-Waterman alignments*. Pac Symp Biocomput. 302-313). Completely annotated proteins were filtered out, whereby 10,130 proteins of unknown function could be grouped into 1,800 clusters.

The obtained sequence clusters were compared to the *Drosophila melanogaster* proteins contained in the database Flybase (Berkeley Drosophila Genome Project; <http://www.fruitfly.org>), and annotated clusters were removed. Non-annotated protein clusters, conserved in both *C. elegans* and *D. melanogaster*, were saved to a worm/fly data set, which was used in a BLAST procedure (<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html>) against the Celera Human Genome Database (<http://www.celera.com>). Overlapping fragments were assembled to, as close as possible, full-length proteins using the PHRAP software, developed at the University of Washington (http://www.genome.washington.edu/UWGC/analysis_tools/phrap.htm). A group of homologous proteins ("Protein Cluster II") with unknown function was chosen for further studies.

EST databases provided by the EMBL (<http://www.embl.org/Services/index.html>) were used to check whether the human proteins in Cluster II were expressed, in order to identify putative pseudogenes. One putative pseudogene was identified and excluded.

EXAMPLE 2: Analyses of Protein Cluster II

(a) Alignment

The human part of Protein Cluster II comprises polypeptides encoded by the nucleic acid sequences shown as SEQ ID NOS: 1, 3 and 5. The sequence shown as SEQ ID NO: 5 was identified as the gene annotated as *Homo sapiens* core1 UDP-galactose:N-acetylgalactosamine-alpha-R beta 1,3-galactosyltransferase (C1GALT1; GenBank Accession No. AF155582; see also International Patent Application No. WO 99/65712.)

An alignment of the human polypeptides included in Protein Cluster II (SEQ ID NOS: 2, 4, and 6), using the ClustalW multiple alignment software (Thompson et al. (1994) Nucleic Acid Research 22: 4673-4680) is shown in Table I. The alignment showed a high degree of conservation in a distinct part of the protein cluster II, indicating the presence of a novel domain (see positions marked with stars in Table I).

The sequence shown as SEQ ID NO: 6 is identical to positions 53-363 of the polypeptide encoded by the C1GALT1 gene mentioned above.

(b) HMM-Pfam

A HMM-Pfam search was performed on the human family members. Pfam is a large collection of protein families and domains. Pfam contains multiple protein alignments and profile-HMMs (Profile Hidden Markov Models) of these families. Profile-HMMs can be used to do sensitive database searching using statistical descriptions of a sequence family's consensus. Pfam is available on the WWW at <http://pfam.wustl.edu>; <http://www.sanger.ac.uk/Software/Pfam>; and <http://www.cgr.ki.se/Pfam>. The latest

version (4.3) of Pfam contains 1815 families. These Pfam families match 63% of proteins in SWISS-PROT 37 and TrEMBL 9. For references to Pfam, see Bateman et al. (2000) *The Pfam protein families database*. Nucleic Acids Res. 28:263-266; Sonnhammer et al. (1998) *Pfam: Multiple Sequence Alignments and HMM-Profiles of Protein Domains*. Nucleic Acids Research, 26:322-325; Sonnhammer et al. (1997) *Pfam: a Comprehensive Database of Protein Domain Families Based on Seed Alignments*. Proteins 28:405-420.

The HMM-Pfam search indicated that no previously known domains could be identified in Protein Cluster II, with exception for a weak homology to Galactosyltransferase (Pfam Accession No. PF01762; see also Kolbinger et al. (1998) J. Biol. Chem. 273: 433-440).

A Pfam-B search revealed identity to the Pfam-B 7357 domain (Pfam Accession No. PB007357). Pfam-B domains are generated automatically from an alignment taken from the database ProDom 2000.1 (<http://www.linux.toulouse.inra.fr/prodom>) subtracting sequence segments already covered by Pfam-A. The ProDom database has been designed as tool to help analyze domain arrangements of proteins and protein families (Corpet et al. (1999) Nucleic Acid Research 27: 263-267). Pfam-B domains are curated manually at the Sanger Centre, UK, to become Pfam-A domains.

(c) TM-HMM

The human proteins in Cluster II were analyzed using the TM-HMM tool available e.g. at <http://www.cbs.dtu.dk/services/TMHMM-1.0>. TM-HMM is a method to model and predict the location and orientation of alpha helices in membrane-spanning proteins (Sonnhammer et al. (1998) *A hidden Markov model for predicting transmembrane helices in protein sequences*. ISMB 6:175-182). No transmembrane regions were identified.



(d) Analysis of non-human orthologs

The *Caenorhabditis elegans* genome includes eight genes encoding proteins within Protein Cluster II, of which the closest ancestor in evolution, a sequence included the *C. elegans* cosmid C38H2.2 (GenBank Accession No. Z35461) and annotated as UDP-galactose:N-acetylgalactosamine- α -R beta 1,3-galactosyltransferase mRNA (GenBank Accession No. AF269063) is 55%, 54%, and 42% identical to the three identified human proteins shown as SEQ ID NOS: 2, 4 and 6, respectively. (See also: *Genome sequence of the nematode C. elegans: a platform for investigating biology*; The *C. elegans* Sequencing Consortium. Science (1998) 282:2012-2018. Published errata appear in Science (1999) 283:35; 283:2103; and 285:1493.)

The *Drosophila melanogaster* genome comprises 10 genes belonging to Protein Cluster II, of which the closest relative "CG9520" (GenBank Accession No. AE003623; see also Adams et al. (2000) *The genome sequence of Drosophila melanogaster*; Science 287:2185-2195) is 42% identical to the human protein set.

No counterparts to Protein Cluster II in *Saccharomyces cerevisiae* were identified.

EXAMPLE 3: Expression analysis

The tissue distribution of the human genes was studied using the Incyte LifeSeq® database (<http://www.incyte.com>). The nucleic acid molecules shown as SEQ ID NO: 1, 3 and 5 were found to be expressed primarily in germ cells and in the nervous system. Therefore, the said nucleic acid molecules shown as SEQ ID NO: 1, 3 and 5 and the polypeptides shown as SEQ ID NO: 2, 4 and 6 are proposed to be useful for differential identification of the tissue(s) or cell types(s) present in a biological sample and for diagnosis of diseases and disorders, including disorders of the central nervous system.

EXAMPLE 4: Multiple Tissue Northern blotting

Multiple Tissue Northern blotting (MTN) is performed to make a more thorough analysis of the expression profiles of the proteins in Cluster II. Multiple Tissue Northern (MTN™) Blots (<http://www.clontech.com/mtn>) are pre-made Northern blots featuring Premium Poly A+ RNA from a variety of different human, mouse, or rat tissues. MTN Blots can be used to analyze size and relative abundance of transcripts in different tissues. MTN Blots can also be used to investigate gene families and alternate splice forms and to assess cross species homology.

EXAMPLE 5: Expressing profiling using microarrays

Microarrays consist of a highly ordered matrix of thousands of different DNA sequences that can be used to measure DNA and RNA variation in applications that include gene expression profiling, comparative genomics and genotyping (For recent reviews, see e.g.: Harrington et al. (2000) *Monitoring gene expression using DNA microarrays*. Curr. Opin. Microbiol. 3(3): 285-291; or Duggan et al. (1999) *Expression profiling using cDNA Microarrays*. Nature Genetics Supplement 21:10-14).

The expression pattern of the proteins in Cluster II can be analyzed using GeneChip® expression arrays (http://www.affymetrix.com/products/app_exp.html). Briefly, mRNAs are extracted from various tissues. They are reverse transcribed using a T7-tagged oligo-dT primer and double-stranded cDNAs are generated. These cDNAs are then amplified and labeled using In Vitro Transcription (IVT) with T7 RNA polymerase and biotinylated nucleotides. The populations of cRNAs obtained are purified and fragmented by heat to produce a distribution of RNA fragment sizes from approximately 35 to 200 bases. GeneChip® expression arrays are hybridized with the samples. The arrays are washed and stained. The cartridges are scanned using a confocal scanner and the images are analyzed with the GeneChip 3.1 software (Affymetrix).

EXAMPLE 6: Identification of polypeptides binding to Protein Cluster II

In order to assay for proteins interacting with Protein Cluster II, the two-hybrid screening method can be used. The two-hybrid method, first described by Fields & Song (1989) *Nature* 340:245-247, is a yeast-based genetic assay to detect protein-protein interactions *in vivo*. The method enables not only identification of interacting proteins, but also results in the immediate availability of the cloned genes for these proteins.

The two-hybrid method can be used to determine if two known proteins (i.e. proteins for which the corresponding genes have been previously cloned) interact. Another important application of the two-hybrid method is to identify previously unknown proteins that interact with a target protein by screening a two-hybrid library. For reviews, see e.g.: Chien et al. (1991) *The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest*. *Proc. Natl. Acad. Sci. U.S.A.* 88:9578-9582; Bartel PL, Fields (1995) *Analyzing protein-protein interactions using two-hybrid system*. *Methods Enzymol.* 254:241-263; or Wallach et al. (1998) *The yeast two-hybrid screening technique and its use in the study of protein-protein interactions in apoptosis*. *Curr. Opin. Immunol.* 10(2): 131-136. See also <http://www.clontech.com/matchmaker>.

The two-hybrid method uses the restoration of transcriptional activation to indicate the interaction between two proteins. Central to this technique is the fact that many eukaryotic transcriptional activators consist of two physically discrete modular domains: the DNA-binding domain (DNA-BD) that binds to a specific promoter sequence and the activation domain (AD) that directs the RNA polymerase II complex to transcribe the gene downstream of the DNA binding site. The DNA-BD vector is used to generate a fusion of the DNA-BD and a bait protein X, and the AD vector is used to generate a fusion of the AD and another protein Y. An entire library of hybrids with the AD can also be constructed to search for new or unknown proteins that interact with the bait protein. When interaction occurs between the bait protein X and a



candidate protein Y, the two functional domains, responsible for DNA binding and activation, are tethered, resulting in functional restoration of transcriptional activation. The two hybrids are cotransformed into a yeast host strain harboring reporter genes containing appropriate upstream binding sites; expression of the reporter genes then indicates interaction between a candidate protein and the target protein.

EXAMPLE 7: Full-length cloning of Cluster II genes

The polymerase chain reaction (PCR), which is a well known procedure for *in vitro* enzymatic amplification of a specific DNA segment, can be used for direct cloning of Protein Cluster II genes. Tissue cDNA can be amplified by PCR and cloned into an appropriate plasmid and sequenced. For reviews, see e.g. Hooft van Huijsduijnen (1998) *PCR-assisted cDNA cloning: a guided tour of the minefield*. Biotechniques 24:390-392; Lenstra (1995) *The applications of the polymerase chain reaction in the life sciences*. Cellular & Molecular Biology 41:603-614; or Rashtchian (1995) *Novel methods for cloning and engineering genes using the polymerase chain reaction*. Current Opinion in Biotechnology 6:30-36. Various methods for generating suitable ends to facilitate the direct cloning of PCR products are given e.g. in Ausubel et al. *supra* (section 15.7).

In an alternative approach to isolate a cDNA clone encoding a full length protein of Protein Cluster II, a DNA fragment corresponding to a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5 or 7, or a portion thereof, can be used as a probe for hybridization screening of a phage cDNA library. The DNA fragment is amplified by the polymerase chain reaction (PCR) method. The primers are preferably 10 to 25 nucleotides in length and are determined by procedures well known to those skilled in the art. A lambda phage library containing cDNAs cloned into lambda phage-vectors is plated on agar plates with *E. coli* host cells, and grown. Phage plaques are transferred to nylon membranes, which are hybridized with a DNA probe prepared as described above. Positive colonies are isolated from the plates. Plasmids containing cDNA are rescued from the isolated phages by standard methods. Plasmid DNA is isolated from the clones. The size of the insert is determined by digesting the plasmid

with appropriate restriction enzymes. The sequence of the entire insert is determined by automated sequencing of the plasmids.

EXAMPLE 8: Recombinant expression of proteins in eukaryotic host cells

To produce proteins of Cluster II, a polypeptide-encoding nucleic acid molecule is expressed in a suitable host cell using a suitable expression vector and standard genetic engineering techniques. For example, the polypeptide-encoding sequence is subcloned into a commercial expression vector and transfected into mammalian, e.g. Chinese Hamster Ovary (CHO), cells using a standard transfection reagent. Cells stably expressing a protein are selected. Optionally, the protein may be purified from the cells using standard chromatographic techniques. To facilitate purification, antisera is raised against one or more synthetic peptide sequences that correspond to portions of the amino acid sequence, and the antisera is used to affinity purify the protein.

EXAMPLE 9: Determination of gene function

Methods are known in the art for elucidating the biological function or mode of action of individual genes. For instance, RNA interference (RNAi) offers a way of specifically and potentially inactivating a cloned gene, and is proving a powerful tool for investigating gene function. For reviews, see e.g. Fire (1999) *RNA-triggered gene silencing*. Trends in Genetics 15:358-363; or Kuwabara & Coulson (2000) *RNAi-prospects for a general technique for determining gene function*. Parasitology Today 16:347-349. When double-stranded RNA (dsRNA) corresponding to a sense and antisense sequence of an endogenous mRNA is introduced into a cell, the cognate mRNA is degraded and the gene is silenced. This type of posttranscriptional gene silencing (PTGS) was first discovered in *C. elegans* (Fire et al., (1998) Nature 391:806-811). RNA interference has recently been used for targeting nearly 90% of predicted genes on *C. elegans* chromosome I (Fraser et al. (2000) Nature 408: 325-330) and 96% of predicted genes on *C. elegans* chromosome III (Gönczy et al. (2000) Nature 408:331-336).

TABLE I

Alignment of polypeptides in Protein Cluster II: "*" = identical or conserved residues in all sequences in the alignment.

```

SEQ_ID_NO_4 -----
SEQ_ID_NO_6 -----
SEQ_ID_NO_2 MTENSLSEMASKSWLNFLTFLYGSAIGFILFSQLLSILLGEEGDTQTNVLHNDPHARHSD 60

SEQ_ID_NO_4 -----NTGVTDKLYQKMILCWIMTGPQNLEKKIRRIRDTWA 37
SEQ_ID_NO_6 DNGQNHLEGQMNFNADSSQHKDENTDIAENLYQKVRILCWVMTGPQNLEKKAKHVKATWA 60
SEQ_ID_NO_2 DNGQNHLEGGQMNFNADSSQRKDENTEIAENLYXQVKILCWVMTGSQNLQKKAKHVKATWA 120
                **      **      ***** ** ** **      ***

SEQ_ID_NO_4 QGCNKALFMSSKENKDFSTVGLHTKEDRNQLSWKIVKAFLYAHDHYLEYMDWFMKADDDI 97
SEQ_ID_NO_6 QRCNKVLFMSSEENKDFPAVGLKTKEGRDQLYWKTIKAFQYVHEHYLEDADWFLKADDDT 120
SEQ_ID_NO_2 QRCLKVFFMSSEENKDFRAVGLKTKAGRDELYWKTINLF----- 159
      * * *      ***** *****      ** * * *      *

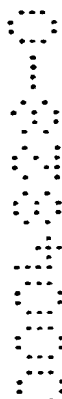
SEQ_ID_NO_4 CIYITLDNLKWLNTNYPDESTYFGKRFKHCRKQDYMTGGAGYVLSKE----- 145
SEQ_ID_NO_6 --YVILDNLRLWLLSKYDPEEPIYFGRRFKPYVKQGYMSGGAGYVLSKEALKRFVDAFKTD 178
SEQ_ID_NO_2 -----

SEQ_ID_NO_4 -----
SEQ_ID_NO_6 KCTHSSSIEDLALGRCMEIMNVEAGDSRDTIGKETFHPFVPEHHLIKGYLPRTFWYWNYN 238
SEQ_ID_NO_2 -----

SEQ_ID_NO_4 -----
SEQ_ID_NO_6 YPPVEGPGCCSDLAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPTLPERILKEISQA 298
SEQ_ID_NO_2 -----

SEQ_ID_NO_4 -----
SEQ_ID_NO_6 NKNEDETKVKGNGP 311
SEQ_ID_NO_2 -----

```



CLAIMS

1. An isolated nucleic acid molecule selected from:
 - (a) nucleic acid molecules comprising a nucleotide sequence as shown in SEQ ID NO: 1 or 3;
 - (b) nucleic acid molecules comprising a nucleotide sequence capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence complementary to the polypeptide coding region of a nucleic acid molecule as defined in (a); and
 - (c) nucleic acid molecules comprising a nucleic acid sequence which is degenerate as a result of the genetic code to a nucleotide sequence as defined in (a) or (b).
2. An isolated polypeptide encoded by the nucleic acid molecule according to claim 1.
3. The isolated polypeptide according to claim 2 having an amino acid sequence shown as SEQ ID NO: 2 or 4 in the Sequence Listing
4. A vector harboring the nucleic acid molecule according to claim 1.
5. A replicable expression vector which carries and is capable of mediating the expression of a nucleotide sequence according to claim 1.
6. A cultured host cell harboring a vector according to claim 4 or 5.
7. A process for production of a polypeptide, comprising culturing a host cell according to claim 6 under conditions whereby said polypeptide is produced, and recovering said polypeptide.
8. A method for identifying an agent capable of modulating a nucleic acid molecule according to claim 1, comprising
 - (i) providing a cell comprising the said nucleic acid molecule;

- (ii) contacting said cell with a candidate agent; and
- (iii) monitoring said cell for an effect that is not present in the absence of said candidate agent.

9
1
2
3
4
5
6
7
8
9
0

SEQUENCE LISTING

<110> Pharmacia AB

<120> Protein Cluster II

<130> 00404

<160> 6

<170> PatentIn version 3.0

<210> 1

<211> 505

<212> DNA

<213> human

<220>

<221> CDS

<222> (21)..(497)

<220>

<221> misc_feature

<222> (297)..()

<223> n=a, g, c or t

<400> 1

gcaagaggga gccacggccg atg aca gaa aat tca ctt tcc gag atg gcc tct	53
Met Thr Glu Asn Ser Leu Ser Glu Met Ala Ser	
1 5 10	
aaa tcc tgg ctg aat ttt tta acc ttc ctc tat gga tgc gca ata ggg	101
Lys Ser Trp Leu Asn Phe Leu Thr Phe Leu Tyr Gly Ser Ala Ile Gly	
15 20 25	
ttt att tta ttt tct cag cta ctt agt att ttg ttg gga gaa gag ggt	149
Phe Ile Leu Phe Ser Gln Leu Leu Ser Ile Leu Leu Gly Glu Glu Gly	
30 35 40	
gac acc cag act aat gtt ctt cat aat gat cct cat gcg agg cat tca	197
Asp Thr Gln Thr Asn Val Leu His Asn Asp Pro His Ala Arg His Ser	
45 50 55	
gat gat aat gga cag aat cat cta gga gga caa atg aac ttc aat gca	245
Asp Asp Asn Gly Gln Asn His Leu Gly Gly Gln Met Asn Phe Asn Ala	
60 65 70 75	
gat tct agc caa cgt aaa gat gag aac aca gaa atc gct gaa aac ctc	293
Asp Ser Ser Gln Arg Lys Asp Glu Asn Thr Glu Ile Ala Glu Asn Leu	
80 85 90	
tat nag caa gtt aaa att ctt tgc tgg gtt atg aca ggc tct caa aac	341
Tyr Xaa Gln Val Lys Ile Leu Cys Trp Val Met Thr Gly Ser Gln Asn	
95 100 105	
cta cag aaa aag gcc aaa cat gtc aaa gct aca tgg gcc cag cgt tgt	389
Leu Gln Lys Lys Ala Lys His Val Lys Ala Thr Trp Ala Gln Arg Cys	
110 115 120	

cta aaa gta ttt ttt atg agt tca gaa gaa aat aaa gac ttc cgt gct 437
 Leu Lys Val Phe Phe Met Ser Ser Glu Glu Asn Lys Asp Phe Arg Ala
 125 130 135

gtg gga ttg aaa acc aaa gca ggc aga gat gag cta tac tgg aaa aca 485
 Val Gly Leu Lys Thr Lys Ala Gly Arg Asp Glu Leu Tyr Trp Lys Thr
 140 145 150 155

att aac ctt ttc agtatggt 505
 Ile Asn Leu Phe

<210> 2
 <211> 159
 <212> PRT
 <213> human

<400> 2
 Met Thr Glu Asn Ser Leu Ser Glu Met Ala Ser Lys Ser Trp Leu Asn
 1 5 10 15

Phe Leu Thr Phe Leu Tyr Gly Ser Ala Ile Gly Phe Ile Leu Phe Ser
 20 25 30

Gln Leu Leu Ser Ile Leu Leu Gly Glu Glu Gly Asp Thr Gln Thr Asn
 35 40 45

Val Leu His Asn Asp Pro His Ala Arg His Ser Asp Asp Asn Gly Gln
 50 55 60

Asn His Leu Gly Gly Gln Met Asn Phe Asn Ala Asp Ser Ser Gln Arg
 65 70 75 80

Lys Asp Glu Asn Thr Glu Ile Ala Glu Asn Leu Tyr Xaa Gln Val Lys
 85 90 95

Ile Leu Cys Trp Val Met Thr Gly Ser Gln Asn Leu Gln Lys Lys Ala
 100 105 110

Lys His Val Lys Ala Thr Trp Ala Gln Arg Cys Leu Lys Val Phe Phe
 115 120 125

Met Ser Ser Glu Glu Asn Lys Asp Phe Arg Ala Val Gly Leu Lys Thr
 130 135 140

Lys Ala Gly Arg Asp Glu Leu Tyr Trp Lys Thr Ile Asn Leu Phe
 145 150 155

<210> 3
 <211> 489
 <212> DNA
 <213> human

<220>
 <221> CDS
 <222> (55)..(489)

<400> 3
catctaaaaa gactgatgaa gttgattgca aatgctagtc atcataaata ccag aac Asn 57
1

aca ggt gtc act gac aaa ctc tat caa aag atg aaa att ctt tgc tgg 105
Thr Gly Val Thr Asp Lys Leu Tyr Gln Lys Met Lys Ile Leu Cys Trp
5 10 15

att atg aca gga cct caa aat cta gaa aaa aag atc aga cgc atc aga 153
Ile Met Thr Gly Pro Gln Asn Leu Glu Lys Lys Ile Arg Arg Ile Arg
20 25 30

gat aca tgg gcc cag ggt tgc aat aaa gcg ttg ttt atg agc tca aaa 201
Asp Thr Trp Ala Gln Gly Cys Asn Lys Ala Leu Phe Met Ser Ser Lys
35 40 45

gaa aat aaa gac ttc tct act gtg gga tta cac acc aaa gaa gac aga 249
Glu Asn Lys Asp Phe Ser Thr Val Gly Leu His Thr Lys Glu Asp Arg
50 55 60 65

aac caa ctg tcc tgg aaa ata gtt aaa gct ttt cta tat gct cat gac 297
Asn Gln Leu Ser Trp Lys Ile Val Lys Ala Phe Leu Tyr Ala His Asp
70 75 80

cat tat ctg gaa tac atg gat tgg ttc atg aaa gca gat gat gat ata 345
His Tyr Leu Glu Tyr Met Asp Trp Phe Met Lys Ala Asp Asp Asp Ile
85 90 95

tgt ata tat atc aca ttg gac aac ttg aaa tgg ctt ctc aca aac tat 393
Cys Ile Tyr Ile Thr Leu Asp Asn Leu Lys Trp Leu Leu Thr Asn Tyr
100 105 110

aac cct gat gaa tcc act tac ttt ggg aaa aga ttt aag cac tgc aga 441
Asn Pro Asp Glu Ser Thr Tyr Phe Gly Lys Arg Phe Lys His Cys Arg
115 120 125

aaa cag gac tac atg act gga gga gca gga tat gta ctg agc aaa gaa 489
Lys Gln Asp Tyr Met Thr Gly Gly Ala Gly Tyr Val Leu Ser Lys Glu
130 135 140 145

<210> 4
<211> 145
<212> PRT
<213> human

<400> 4
Asn Thr Gly Val Thr Asp Lys Leu Tyr Gln Lys Met Lys Ile Leu Cys
1 5 10 15

Trp Ile Met Thr Gly Pro Gln Asn Leu Glu Lys Lys Ile Arg Arg Ile
20 25 30

Arg Asp Thr Trp Ala Gln Gly Cys Asn Lys Ala Leu Phe Met Ser Ser
35 40 45

Lys Glu Asn Lys Asp Phe Ser Thr Val Gly Leu His Thr Lys Glu Asp
50 55 60

Arg Asn Gln Leu Ser Trp Lys Ile Val Lys Ala Phe Leu Tyr Ala His
 65 70 75 80
 Asp His Tyr Leu Glu Tyr Met Asp Trp Phe Met Lys Ala Asp Asp Asp
 85 90 95
 Ile Cys Ile Tyr Ile Thr Leu Asp Asn Leu Lys Trp Leu Leu Thr Asn
 100 105 110
 Tyr Asn Pro Asp Glu Ser Thr Tyr Phe Gly Lys Arg Phe Lys His Cys
 115 120 125
 Arg Lys Gln Asp Tyr Met Thr Gly Gly Ala Gly Tyr Val Leu Ser Lys
 130 135 140
 Glu
 145

<210> 5
 <211> 1560
 <212> DNA
 <213> human

<220>
 <221> CDS
 <222> (2)..(934)

<400> 5
 a gat aat gga cag aat cat cta gaa gga caa atg aac ttc aat gca gat 49
 Asp Asn Gly Gln Asn His Leu Glu Gly Gln Met Asn Phe Asn Ala Asp
 1 5 10 15
 tct agc caa cat aaa gat gag aac aca gac att gct gaa aac ctc tat 97
 Ser Ser Gln His Lys Asp Glu Asn Thr Asp Ile Ala Glu Asn Leu Tyr
 20 25 30
 cag aaa gtt aga att ctt tgc tgg gtt atg acc ggc cct caa aac cta 145
 Gln Lys Val Arg Ile Leu Cys Trp Val Met Thr Gly Pro Gln Asn Leu
 35 40 45
 gag aaa aag gcc aaa cac gtc aaa gct act tgg gcc cag cgt tgt aac 193
 Glu Lys Lys Ala Lys His Val Lys Ala Thr Trp Ala Gln Arg Cys Asn
 50 55 60
 aaa gtg ttg ttt atg agt tca gaa gaa aat aaa gac ttc cct gct gtg 241
 Lys Val Leu Phe Met Ser Ser Glu Glu Asn Lys Asp Phe Pro Ala Val
 65 70 75 80
 gga ctg aaa acc aaa gaa ggc aga gat caa cta tac tgg aaa aca att 289
 Gly Leu Lys Thr Lys Glu Gly Arg Asp Gln Leu Tyr Trp Lys Thr Ile
 85 90 95
 aaa gct ttt cag tat gtt cat gaa cat tat tta caa gat gct gat tgg 337
 Lys Ala Phe Gln Tyr Val His Glu His Tyr Leu Gln Asp Ala Asp Trp
 100 105 110
 ttt ttg aaa gca gat gat gac acg tat gtc ata cta gac aat ttg agg 385
 Phe Leu Lys Ala Asp Asp Asp Thr Tyr Val Ile Leu Asp Asn Leu Arg
 115 120 125

tgg ctt ctt tca aaa tac gac cct gaa gaa ccc att tac ttt ggg aga	433
Trp Leu Leu Ser Lys Tyr Asp Pro Glu Glu Pro Ile Tyr Phe Gly Arg	
130 135 140	
aga ttt aag cct tat gta aag cag ggc tac atg agt gga gga gca gga	481
Arg Phe Lys Pro Tyr Val Lys Gln Gly Tyr Met Ser Gly Gly Ala Gly	
145 150 155 160	
tat gta cta agc aaa gaa gcc ttg aaa aga ttt gtt gat gca ttt aaa	529
Tyr Val Leu Ser Lys Glu Ala Leu Lys Arg Phe Val Asp Ala Phe Lys	
165 170 175	
aca gac aag tgt aca cat agt tcc tcc att gaa gac tta gca ctg ggg	577
Thr Asp Lys Cys Thr His Ser Ser Ser Ile Glu Asp Leu Ala Leu Gly	
180 185 190	
aga tgc atg gaa att atg aat gta gaa gca gga gat tcc aga gat acc	625
Arg Cys Met Glu Ile Met Asn Val Glu Ala Gly Asp Ser Arg Asp Thr	
195 200 205	
att gga aaa gaa act ttt cat ccc ttt gtg cca gaa cac cat tta att	673
Ile Gly Lys Glu Thr Phe His Pro Phe Val Pro Glu His His Leu Ile	
210 215 220	
aaa ggt tat cta cct aga acg ttt tgg tac tgg aat tac aac tat tat	721
Lys Gly Tyr Leu Pro Arg Thr Phe Trp Tyr Trp Asn Tyr Asn Tyr Tyr	
225 230 235 240	
cct cct gta gag ggt cct ggt tgc tgc tct gat ctt gca gtt tct ttt	769
Pro Pro Val Glu Gly Pro Gly Cys Cys Ser Asp Leu Ala Val Ser Phe	
245 250 255	
cac tat gtt gat tct aca acc atg tat gag tta gaa tac ctc gtt tat	817
His Tyr Val Asp Ser Thr Thr Met Tyr Glu Leu Glu Tyr Leu Val Tyr	
260 265 270	
cat ctt cgt cca tat ggt tat tta tac aga tat caa cct acc tta cct	865
His Leu Arg Pro Tyr Gly Tyr Leu Tyr Arg Tyr Gln Pro Thr Leu Pro	
275 280 285	
gaa cgt ata cta aag gaa att agt caa gca aac aaa aat gaa gat aca	913
Glu Arg Ile Leu Lys Glu Ile Ser Gln Ala Asn Lys Asn Glu Asp Thr	
290 295 300	
aaa gtg aag tta gga aat cct tgaaagaaaa tcatgaatga acaaaggtaa	964
Lys Val Lys Leu Gly Asn Pro	
305 310	
tatgtctagc actgcactga aaaaggactt ctgcatttct gacatagaac actggaatcc	1024
cagtgaggaa ttctaagtga acattcctta tagaaacctt tcacatgaat gactataaac	1084
tgaagcttta aatgagctgt gaagtgtgtt aaaatgtgtt ttgatacagt aatatataaa	1144
tatgtctata tatatgagga acttgtgttt tttaaagtgtt ggccaggtag aggaactaga	1204
aaagagattt tgttgctgt tttctgacca tctgtgttat tgctactgag aaactaaaat	1264
agtaaattta ctaaaactac actgcacccat gttagtaata aacagatctg ccttaaagaa	1324

aagaaaattt tagaaagaaa tattgttgct cagtgttggt aatatagctc aagaattgag 1384
 tttatatattg cagtatgcta taaatgatac cccctacca caccacaca cacagttttt 1444
 gtctaatagaa aatgttgctg tgattattta taattggtag tattttctcc agaagaagct 1504
 aaaataagac tggcacttac cctgaagtgc attaataaaa ccacacttta aaatta 1560

<210> 6
 <211> 311
 <212> PRT
 <213> human

<400> 6
 Asp Asn Gly Gln Asn His Leu Glu Gly Gln Met Asn Phe Asn Ala Asp
 1 5 10 15
 Ser Ser Gln His Lys Asp Glu Asn Thr Asp Ile Ala Glu Asn Leu Tyr
 20 25 30
 Gln Lys Val Arg Ile Leu Cys Trp Val Met Thr Gly Pro Gln Asn Leu
 35 40 45
 Glu Lys Lys Ala Lys His Val Lys Ala Thr Trp Ala Gln Arg Cys Asn
 50 55 60
 Lys Val Leu Phe Met Ser Ser Glu Glu Asn Lys Asp Phe Pro Ala Val
 65 70 75 80
 Gly Leu Lys Thr Lys Glu Gly Arg Asp Gln Leu Tyr Trp Lys Thr Ile
 85 90 95
 Lys Ala Phe Gln Tyr Val His Glu His Tyr Leu Gln Asp Ala Asp Trp
 100 105 110
 Phe Leu Lys Ala Asp Asp Asp Thr Tyr Val Ile Leu Asp Asn Leu Arg
 115 120 125
 Trp Leu Leu Ser Lys Tyr Asp Pro Glu Glu Pro Ile Tyr Phe Gly Arg
 130 135 140
 Arg Phe Lys Pro Tyr Val Lys Gln Gly Tyr Met Ser Gly Gly Ala Gly
 145 150 155 160
 Tyr Val Leu Ser Lys Glu Ala Leu Lys Arg Phe Val Asp Ala Phe Lys
 165 170 175
 Thr Asp Lys Cys Thr His Ser Ser Ser Ile Glu Asp Leu Ala Leu Gly
 180 185 190
 Arg Cys Met Glu Ile Met Asn Val Glu Ala Gly Asp Ser Arg Asp Thr
 195 200 205
 Ile Gly Lys Glu Thr Phe His Pro Phe Val Pro Glu His His Leu Ile
 210 215 220

Lys Gly Tyr Leu Pro Arg Thr Phe Trp Tyr Trp Asn Tyr Asn Tyr Tyr
225 230 235 240

Pro Pro Val Glu Gly Pro Gly Cys Cys Ser Asp Leu Ala Val Ser Phe
245 250 255

His Tyr Val Asp Ser Thr Thr Met Tyr Glu Leu Glu Tyr Leu Val Tyr
260 265 270

His Leu Arg Pro Tyr Gly Tyr Leu Tyr Arg Tyr Gln Pro Thr Leu Pro
275 280 285

Glu Arg Ile Leu Lys Glu Ile Ser Gln Ala Asn Lys Asn Glu Asp Thr
290 295 300

Lys Val Lys Leu Gly Asn Pro
305 310

00404-SE

ABSTRACT

The present invention relates to the identification of a human gene family expressed in metabolically relevant tissues. The genes encode a group polypeptides referred to as "Protein Cluster II" which are predicted to be useful in the diagnosis of metabolic diseases, such as obesity and diabetes, as well as in the identification of agents useful in the treatment of the said diseases.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2